Quasispecies: from molecular Darwinism to viral diseases

Esteban Domingo
Severo Ochoa Center for Molecular Biology, CSIC-UAM, Madrid, Spain
Network Center for Biomedical Research in Hepatic and Digestive Diseases (CIBERehd), Barcelona, Spain

Summary. The microbial world offers the most direct evidence of natural selection, the central concept of Darwin’s theory. The evidence of the extraordinary biological diversity on the molecular level, and its parallelism to morphological diversity observed by Darwin, has continued to increase until our days. It affects all types of organisms and the parasites they contain, including viruses. Some experimental designs with RNA viruses have allowed us to dissect the basic processes of biological evolution: genetic variation, competition and selection. The population dynamics that characterizes RNA viruses is known as quasispecies dynamics, which refers to a theory of the origin of life developed four decades ago, which proposes that the first elements equipped with autonomous replication were constructed, approximately 4000 million years ago, from small molecules, equal or similar to the RNA we know today. These primitive molecules were able to evolve thanks to the continuous production of erroneous copies, as we observe in the current RNA viruses. Today, quasispecies dynamics allows viruses to survive in the organisms they parasitize and respond to selective pressures that attempt to stop their multiplication (components of the immune system, drugs, etc.). Among RNA viruses there are well-known human pathogens such as the virus that causes AIDS, the viruses associated to the different forms of hepatitis, and different emerging and reemerging viruses, so the implications of the dynamics of quasispecies for the control of viral diseases are very clear. Research has established that within viral populations there are interactions between components of the same quasispecies, which behave/operate as a “selection unit”. This observation represents a fundamental change with regard to what was thought a few years ago, in the sense that the behavior of viruses is not necessarily predictable by the behavior of the individual genomes that make up a population. This has several theoretical, as well as medical, implications.

Keywords: biological evolution · RNA viruses · quasispecies · quasispecies dynamics · lethal mutagenesis
The contributions of Darwin have been associated fundamentally with biological evolution. Yet, natural selection, the key concept of Darwinism, has been influential in a wide range of phenomena, most of them paramount to the understanding of the universe we live in and even to the origin of life on Earth. The essence of Darwinism can be illustrated by the developments shown in Table 1.

Table 1. The ample influence of Darwinism

- Evidence that evolution did indeed occur and that humans are a product of biological evolution
- Cosmological natural selection
- Origin of life; quasispecies theory
- Brain function; perception, cognition, memory, behavior; evolutionary psychiatry
- Social Darwinism

*See text for references.

Some of them still surrounded by controversy:

First, Darwin provided extensive empirical documentation that evolution did indeed occur and that humans (like any other animals or plants) are a product of biological evolution. This represented a fundamental change of perspective on the nature of human beings and their place in the biosphere. It was a direct blow against religious views of the origin of man. The shock must have been equivalent to the recognition by Nicolaus Copernicus that planet Earth was not the center of the solar system but just one planet among others, in orbit around a star just like many other planets in a universe full of stars. The change in perception of the nature of humans necessarily created a strong reaction from those that defended other, scientifically unfounded notions on the origin of humans and the "purpose" of their existence. In the eyes of a dominant religious establishment that has opposed (or opposes) undertakings such as the dissection of human cadavers, anesthesia, surgery, birth control, in vitro fertilization, or stem-cell research (among others), the departure proposed by Darwin was, no doubt, devastating. Resistance to the theory of evolution extends to the present time, in the form of creationism or its more recent version, intelligent design.

Second, in 1992 the cosmologist Lee Smolin proposed the theory of cosmological natural selection (CNS), applying Darwinian concepts to the origin of the Universe [30,31]. In the CNS, the notion of the "landscape" of the Universe or an "ensemble" of universes was introduced, in analogy with "fitness landscape" or "space of genotypes" in biology. The space of genotypes can be projected into a space of phenotypes. Likewise, in the CNS model, a space of parameters, analogous to the space of phenotypes in biology, includes those parameter values that give rise to universes similar to ours, with long-lived stars and a complex chemistry compatible with life. The fitness function is the average number of black holes generated by each "bounce transition" or "creation" event. CNS constitutes the main current alternative to the multiverse cosmological theory (the existence of an infinite number of universes) based on the anthropic principle (for a general account of the anthropic principle and the multiple universe concept, see [35]). The controversy between these two cosmological theories continues, but the formulation of the CNS in a scientific domain so distant from biology is a demonstration of the far-reaching influence of Darwinian concepts.

Third, on a widely different time-space scale, the understanding of brain function also has been guided by natural selection and evolutionary concepts. Processes such as perception, cognition, and memory have been viewed as events that result from the selection of neuronal contacts, triggered by external stimuli. William Calvin popularized this view of brain function and its links with Darwinism [3].

And fourth, a better known extension of Darwinian concepts to social sciences is social Darwinism, which asserts that competition at different levels (among individuals, organized groups, nations, etc.) drives social evolution in human societies. According to this view, competition leads to "survival of the fittest," a term coined by Herbert Spencer. The development of social Darwinism was influenced not only by the ideas of Darwin and Spencer (in some aspects, the ideas of Spencer predated those of Darwin), but also by Thomas Malthus (who also exerted an important influence on Darwin) and Francis Galton, the founder of eugenics. The multiple facets and impact of social Darwinism, including eugenics and Nazi ideology, are exploited even today to oppose Darwin. Their coverage is beyond the scope of this article.

What is the key concept that unites the Darwinian view of the origin of man, the origin of our universe, and brain function? It is "dynamics." Natural selection (acting on living or non-living objects) implies the generation of multiple versions of a given object and, at the same time, an external (environmental) influence that makes one of the versions prevail over the others. Selection occurs within widely different time scales: millions of years for the predominance of some animal species and fractions of a second for the preferred reinforcement of some neuronal contacts.

Origin of life and quasispecies theory

One of the most influential theories in biology developed during the 20th century was the quasispecies theory, initiated in the work of Manfred Eigen [11] and then developed by Eigen and Schuster [12]. Quasispecies is a theory of the origin of life that seeks to explain the transition from an inanimate world to the first forms of life (self-replicating entities), which is assumed to have consisted of RNA or RNA-like replicons. The theory describes mathematically error-prone replication and the formation of abundant collections of mutants, termed mutant clouds. The generation of diversity allowed the self-organization and adaptability of the primitive replicons that might have populated the Earth some 4000 million years ago, a time known as "the RNA world." A highly schematic and simplified view of the possible course of events is depicted in Fig. 1. Myriads of simple, non-replicative RNAs (or RNA-like oligonucleotides) could have been synthesized under pre-biotic conditions [24]. Among them, a few could have displayed an incipient replicative (self-
copying) capacity, as suggested by the catalytic activities of some present-day ribozymes (RNA enzymes). When RNA levels reached a critical replicative level (center panel in Fig. 1), additional diversification and the acquisition of new functions became possible over extended periods of time, prior to the advent of DNA, that involved specific protein catalysts, compartmentalization, and mechanisms for the generation or capture of energy-rich compounds for metabolic activity. Obviously, other mechanisms for the origin of life have been proposed, highlighting either an incipient metabolism or primitive genetics as the main triggering event (reviewed in [1]). Despite problems in the laboratory reproduction of some of the steps likely involved in the synthesis of self-replicating molecules, our current knowledge of chemistry and biology offers strong arguments that primitive life was created naturally from inert (non-living) building blocks.

In connection with quasispecies theory, an unexpected event occurred in 1978, at the “Winterseminar” held in Klosters (Switzerland), a meeting sponsored annually by the Max Planck Institute and organized by Eigen. Just when the formulation of quasispecies theory had been completed by Eigen and Schuster, Charles Weissmann reported in Klosters that the structure and dynamics of populations of an RNA virus, the bacteriophage Qβ, were precisely as predicted by quasispecies theory [5]. This was the beginning of an extremely influential collaboration between theoretical biophysics and virology that continues today (for an historical account of the development of quasispecies in virology, including a celebration at the Institute for Catalan Studies in November 2008, see [9]). Quasispecies theory represents a link between concepts of information theory and Darwinian natural selection, and it has been remarkably useful in furthering our understanding of genetic systems that replicate with limited copying fidelity.

### Darwin and microbiology

The year 1677 is generally agreed to be the date of the origin of microbiology as a branch of the natural sciences. In that year, Antonie van Leeuwenhoek described his observations of small living creatures visible under the microscope. The results were published as a letter in the Philosophical Transactions of the Royal Society of London [32]. Although the studies of van Leeuwenhoek preceded those of Darwin by almost two centuries, microbiology and infectious diseases were not taken into consideration in the development of evolutionary thinking. Darwin was aware of the work of Louis Pasteur and Robert Koch.
but did not mention infectious disease as possible promoters of evolution (review in [20]).

Cellular microbes and viruses are relevant in evolutionary biology for two reasons. First, they can evolve at very rapid rates; as fast evolvers they shed light on key evolutionary mechanisms, such as competition, selection, and bottleneck events (the colonization of some environment by one or a few individuals separated from a large microbial population). Second, pathogenic microbes are the actors of evolution, modulating host population numbers and favoring selection of humans, animals, and plants that are resistant to decimating diseases. In this scenario, the fastest replicating and evolving elements that have been recognized to date are RNA genetic elements, and particularly RNA viruses. Viruses were not discovered until 1898 [21], 39 years after the publication of Darwin’s *On the Origin of Species*. Decades later, viruses were defined as particles containing either DNA or RNA (but not both) as genetic material and whose replication is totally dependent on a host cell. Viruses have served as the source of defined nucleic acids of different conformations (linear, circular, single-stranded, double-stranded, segmented or unsegmented) for physical, genetic, and biochemical studies. Our understanding of basic life processes, such as replication, transcription (synthesis of RNA from DNA), and translation (protein synthesis), has greatly benefited from the use of viral nucleic acids as templates.

Interestingly, during replication, viruses that have RNA as their genetic material, the RNA viruses, mutate at rates that can be up to a million-fold higher than the rates operating normally during cellular DNA replication (Fig. 2) (reviews in [7,18]). RNA viruses include important animal pathogens, such as human immunodeficiency virus type 1, influenza virus, and the viruses of hepatitis A and C. The study of their evolution has shed light not only on their origin but also on mechanisms of pathogenesis, the latter being intimately related to quasispecies dynamics [7,9,19].

**Microbes and evolution in the 20th century**

The impact of rapid evolution of RNA viruses must be placed in the context of the developments of evolutionary biology that took place in the 20th century. The 1920s witnessed the formulation of the “modern synthesis,” i.e., a unification of the concepts of natural selection and Mendelian genetics, as proposed by Ronald A. Fisher, John B.S. Haldane, and Sewall Wright. The “synthesis” was coherent with a number of key discoveries in molecular genetics, such as that of DNA as the genetic material and the mispairing of bases as a mechanism of mutation, and with the flow of genetic information being DNA → RNA → protein (with the exception of retroviruses and other retroelements that contain reverse transcriptase, an enzyme enabling the synthesis of DNA from RNA). Advances in molecular biology, including the advent of genetic engineering and reverse genetics, reinforced the conviction that Darwinian principles provided the framework to understand evolution. Indeed, there was little question that genetic material underwent variation (by mutation, several forms of recombination, and genome segment or chromosome reassortment) and that there was competition among variant forms as well as selection of the most fit individuals (or collectivity of individuals) under a given set of environmental conditions. These Darwinian principles apply to the biological world in its entirety, including microbes and viruses, and are consistent with the nature of the genetic material and its variations.

**The relevance of quasispecies in viral evolution**

When an RNA virus infects a host organism, a mutant spectrum is generated due to the error-prone replication of the incoming virus (Fig. 3). Organisms can be infected with a single virus particle (in the case that a severe bottleneck occurred during transmission), or by multiple, closely related particles (from a mutant spectrum present in the donor, infected host), or by particles...
from different isolates. In addition, and to complicate matters further, infection may reactivate latent, related or unrelated viruses that were dormant in the organism. In all cases, further replication of the incoming virus can be regarded as a first stage of virus diversification within an infected host. To some extent, each infection in vivo represents the unfolding of a new evolutionary history for a virus. Viruses that we isolate and analyze are the product of successive cycles of transmission, a unique evolutionary history in an individual host, new transmission, etc.

Analyses of the mutant composition of RNA viruses, as well as of at least some DNA viruses, have documented that they replicate as mutant spectra (also termed mutant clouds) rather than as a defined genomic nucleotide sequence (recent reviews in [6,8]), as predicted by quasispecies theory. However, the latter was formulated initially as a deterministic theory involving steady-state mutant distributions of infinite size, in equilibrium. Under the assumption of a single peak fitness landscape, the population is dominated by a master sequence, which is the genome with maximum fitness in the distribution [11,12]. Obviously, mutant spectra of real viruses are not steady-state distributions located in a single peak fitness landscape. Generally, deterministic models are developed first because they place a problem in mathematically solvable terms. At later stages, stochastic components are introduced to render the models more realistic. This and other simplifications introduced in the development of quasispecies theory and its implications (for example, the assumption of a single fitness peak, and the assignment of the same relative fitness value to all components of a mutant spectrum) were a necessity given the limited computational power and tools at the time of the theory’s formulation. However, such simplifications should not be used to diminish the relevance of quasispecies to understand RNA viruses. More recently, extensions of quasispecies theory to finite genome populations in variable environments have been developed [13,29,36], and they further

Table 2. Biological relevance of mutant spectra

- Mutant spectra are reservoirs of genetic and phenotypic virus variants.
- Quasispecies complexity can affect biological behavior (virulence, response to antiviral treatments).
- Interactions among components of a mutant spectrum are established and can be either positive (producing complementation) or negative (producing interference).
- Mutant spectra may include “memory” genomes.
- Viral quasispecies act as a “unit of selection.”

*See text for references.

Fig. 3. Schematic description of a viral quasispecies. Infection by a single viral genome results in the generation of mutant spectra, one of which is represented on the right. Genomes are depicted as horizontal lines, and mutations as symbols on the lines. In real infections, the number of genomes at any given time can be as high as $10^{11}$–$10^{12}$ particles, constituting vast mutant clouds. The consensus sequence (bottom) is the one that includes at each position the residue found most frequently at the corresponding position in the mutant ensemble. The complexity (average number of mutations per genome) and composition of a viral quasispecies are important determinants of the biological behavior of a virus. See text for specific examples and references.
justify it as an adequate theoretical framework for the understanding of RNA virus evolution.

**Quasispecies as reservoirs of variant viruses and as integrated units of selection**

The existence of RNA viruses as mutant spectra has several consequences that affect the interpretation of both virus evolution and viral pathogenesis, as summarized in Table 2.

The main points are the following:

First, mutant spectra are extensive reservoirs of genetic and phenotypic variation. Until recently, the abundance of mutant spectra could be assessed by determining nucleotide sequences of individual molecular or biological clones that compose a viral population. This allowed mutant spectra to be defined at the resolution level of 10–100 clones per sample (reviewed in [6,7]). New deep-sequencing technologies can be used to penetrate the composition of mutant spectra, and to analyze $10^4$–$10^5$ clones per sample, with a resolution of minority mutations present at a frequency of 1% [23]. These new technologies have amply confirmed the existence of complex mutant spectra for all viral isolates analyzed to date. The array of variants confers viruses with the ability to adapt to different environments. Examples relevant to intra-host development of an infection are variants with altered host-cell tropism (allowing viral progeny to infect a new tissue or organ), variants that can escape neutralization by antibodies or cytotoxic T cells (facilitating viral persistence), or variants with decreased sensitivity to antiviral inhibitors administered to limit viral replication.

Second, quasispecies complexity (the amplitude of the mutant spectrum (Fig. 3)) can affect biological behavior (several examples are provided in [6]). This has been best documented in a poliovirus mutant exhibiting a template copying fidelity about five times higher than that of the wild-type virus. The increased fidelity was due to a single amino acid substitution in the viral RNA-dependent RNA polymerase, the enzyme involved in replication of the viral genome. This high-fidelity mutant gave rise to mutant spectra that were narrower (lower average number of mutations per genome) than the wild-type spectrum. Interestingly, the mutant had a defect in adaptability to a complex environment and displayed decreased neuropathology in mice [27,33]. These results demonstrated the biological relevance of high mutation rates in vivo (they can no longer be regarded as a mere consequence of rapid replication), illustrated the biological relevance of mutant spectra, and introduced a new concept of virulence and attenuation as traits that can be determined by the complexity of mutant spectra [34].

Third, mutant spectra are not simply collections of mutants acting independently. They constitute integrated units of selection. This is supported by different lines of evidence, involving either the analysis of selected genomes, the effect of perturbations introduced into a viral quasispecies, or the presence of memory genomes in viral populations. The main observations can be summarized as follows. When a quasispecies was reconstructed with low frequencies of mutants resistant to a monoclonal antibody, a mutant cloud, not an individual variant, was selected by the antibody [25]. This result implied that collections of mutants can be selected, even though no obvious interactions among the individual mutants used to reconstruct the quasispecies were expected. In another approach, a highly mutagenized mutant spectrum interfered with replication-competent standard genomes present in the same quasispecies [16]. This was one of the observations that supported lethal defection as a model of virus extinction by enhanced mutagenesis [17]. According to the lethal defection model, which is currently the object of considerable theoretical and experimental research, a class of defective genomes that interfere with the replication of a standard virus can contribute to virus extinction, when the virus replicates under conditions of enhanced mutagenesis. Indeed, it has been documented that individual viral mutants can either complement or interfere with the replication of other mutants or of standard genomes with which they coexist in the same mutant spectrum [4,26]. In several early studies it was observed that the fitness of biological clones isolated from a viral quasispecies was less than the fitness of the average, parental population [5,10]. This is consistent with the occurrence of complementation among components of a quasispecies when replicating under a basal mutation rate. Complementation becomes interference when mutation rates increase and defective genomes become more abundant in the population.

And fourth, the relevance of the mutant spectrum as a whole for virus behavior is also illustrated by the presence of "memory" in viral populations. Indeed, evolving quasispecies are endowed with a molecular memory, consisting of minority genomes reflecting those that were dominant at a previous phase of the same evolutionary lineage [28]; review in [2]. Since population bottlenecks erase memory, the latter is a property of the ensemble of genomes that compose the entire quasispecies. Memory may favor adaptation of a virus population to an environmental change previously experienced by the same lineage at an earlier phase of its evolution.

The first and second features document the multiple biological implications of mutant spectra regarding adaptability and interaction with the environment, in particular as a factor in viral pathogenesis. Features summarized in the third point demonstrate that there are internal interactions among components of the same replicating mutant spectrum, exerted through trans-active gene products. Quasispecies behavior is influenced by such intra-mutant spectrum interactions. The fourth feature shows that quasispecies dynamics provides a means to respond in a history-dependent manner, a property typical of complex adaptive systems ([14,15; discussed in [28]).

Therefore, viral quasispecies cannot be regarded as independently acting collections of mutants that reach a mutation-selection equilibrium, as in classical Wright-Fisher formulations of population genetics. Rather, their behavior is collective and difficult to predict from the behavior of the individual components.

**Viruses as models of biological complexity**

The application of quasispecies to the understanding of viral population dynamics introduced the concepts of complexity to
the field of virology and, to a certain extent, also to Darwinian variation-competition-selection principles. Complex behavior is shared by a number of physical, chemical, and biological systems and processes. Complexity governs disparate operations such as brain function, climate evolution, the spread of epidemics, the activity of financial markets, the immune system, the response to therapeutic interventions, and the motion of plasmas (gases of charged particles) (reviewed in [22]). They share collective patterns of interactions, and the resulting behavior cannot be anticipated from that of the individual elements comprising the ensemble.

Complexity in biological systems can be regarded as an extension of the mechanisms by which natural selection operates. In this new view, the targets of selection are not “individual objects” subjected to variation but “a group of interacting objects” subjected to variation. Occasionally, complex behavior has been erroneously equated with the complexity of a biological object and taken as evidence of some mysterious, higher-order interventions in the course of natural events. Indeed, the so-called principle of “irreducible complexity” has been used as evidence of an intelligent design. This is obviously a biased and unscientific interpretation. Whether we refer to the complexity of an object or to complexity in behavior, the available evidence indicates that either one is the result of the same natural laws of physics, chemistry, and biology that have shaped (and are shaping) the physical and biological worlds. In this understanding of life processes, the Darwinian concept of natural selection occupies a central role.

To finish this brief account of viruses as Darwinian systems, here I reproduce the last sentence in Calvin’s book on brain function [3], which illustrates the fascinating potential of matter to organize itself to give rise to mental activity, achieved following the natural rules of physics, chemistry, and biology: “The Darwinian principles that shaped life on earth over billions of years, that daily reshape the immune system in our bodies, have again blossomed in human heads on an even more accelerated time scale. In much the manner that life itself unfolded, our mental life is progressively enriched enabling each of us to create our own world. To paraphrase Charles Darwin, there is grandeur in this view of mind.”

Acknowledgements

Work in the author’s laboratory is supported by grants BFU2008-02816/BMC from the Ministry of Science and Innovation, Intramural de Frontera Project CSIC 200820FO191, and Ramón Areces Foundation. CIBERehd (Network Center for Biomedical Research in Hepatic and Digestive Diseases) is funded by the Carlos III Health Institute.

References

About the author

Esteban Domingo graduated in Chemical Sciences and obtained his Ph.D. in Biochemistry at the University of Barcelona, both with distinctions. He carried out postdoctoral research at the University of California-Irvine, where he studied bacterial RNA polymerases and transcription of viral DNAs under the supervision of Robert C. Warner. Then, he moved to the University of Zurich, where, working with Charles Weissmann on the Qβ bacteriophage, he determined the first error rate for an RNA virus and discovered the dynamics of RNA virus quasispecies, which he continued studying with animal viruses in Madrid. He has directed the Severo Ochoa Center for Molecular Biology (CSIC-CBMSO) and the Animal Health Research Center (CISA-INIA). He is currently research professor at the CBMSO. He has published more than 250 papers on viral evolution and new strategies for the control of viral diseases. He is scientific founder of the Center for Astrobiology (CSIC-CAB) and the Institute Para Limes (The Netherlands), and is president of the Spanish Society of Virology. He has a great interest in transdisciplinary science, and is the author and editor of several books, a member of the European Molecular Biology Organization (EMBO), of the European Academy, and has honorary doctorates (honoris causa) from the Universities of Liege and Bern. He is one of the most cited Spanish researchers.